USING MODERNIZED STORET TO CREATE A STATE-WIDE DATA CLEARINGHOUSE IN IOWA

Mary P. Skopec[1] and Joost Korpel[1]

[1] Iowa Department of Natural Resources, Iowa Geological Survey
109 Trowbridge Hall, Iowa City, Iowa, 52242-1319.

Biographical Sketch of Authors

Mary Skopec is the Section Supervisor for the Water Monitoring Program at the Iowa Department of Natural Resources, Geological Survey Bureau (GSB). Mary earned her B.S. and M.A. degrees in geography from the University of Iowa, and in 1999 she completed her Ph.D. in environmental science. Mary has worked at the Geological Survey since 1991, and during that time she has been involved been involved in a variety of water quality projects including the development of a statewide pesticide database to track pesticide occurrences in Iowa's water resources.

Joost Korpel is an information technology specialist with Iowa's Water Monitoring Program. He received his master's degree in geology from the University of Iowa, and has worked in Iowa state government for 19 years in various roles including systems analyst, research geologist, geographic information specialist, and IT network administrator. Joost provides expertise in both geology and IT data management (including the IOWATER database and Iowa's STORET database) to the Water Monitoring Program.

Abstract

Iowa's ambient water quality program is a diverse program responsible for the collection and analysis of information on the state's rivers, lakes, groundwater and wetlands. Physical, chemical, biological and habitat data are administered as part of this program. To handle this complex data set, the Iowa Department of Natural Resources (IDNR) is using the Environmental Protection Agency (EPA) database called STORET (STOrage and RETrieval). Not only does STORET allow the DNR to comprehensively handle its water monitoring data, the use of STORET also ensures compatibility with data from other state and federal agencies. It is envisioned that STORET will be a "one-stop shopping" location for all the state's water monitoring data. This would include the coordination of data from other entities such as municipalities, citizen monitoring groups, universities and others into the STORET data management system. The public retrieval of data currently includes a Web-based interface that provides an interactive mapping application. This allows a user to click on a station or set of stations to view the data, display the data in graphical form, or see statistical summaries. Users are able to download data in a variety of formats, including flat-files and Excel files.

Introduction

In 1999, the state of Iowa initiated a new water-monitoring program to measure ambient conditions of the state's water resources. Prior to this time, the water-monitoring program was very limited, consisting of sixteen surface water stations sampled monthly for nutrients and water chemistry and ninety groundwater wells monitored annually on a rotational basis. Funding for the old program came from a combination of state general fund and U.S. Environmental Protection Agency grants with an annual operating budget of approximately $300,000 (evenly split between state and federal funding). State funding for the new program has increased to its current level of $2.5 million in fiscal year 2002, which represents a significant commitment toward monitoring water quality in the state.

Iowa's ambient water quality program is a diverse program responsible for the collection and analysis of information on the state's rivers, lakes, groundwater and wetlands. Physical, chemical, biological and habitat data are administered as part of this program. Additionally, meetings with stakeholders to determine the scope and direction of the new program indicated that coordination of existing data from other public and private agencies was a high priority. Most stakeholders were concerned that much of data collected in the state was not available to the public or appropriate decision-makers. The Iowa Department of Natural Resources (IDNR) also acknowledged the value in coordinating and migrating long-term datasets into a comprehensive data system given the lack of historical monitoring in the state. The development of a central data warehouse or clearinghouse required the use of a powerful database engine that could accommodate and document the various ways data had been collected and analyzed by the various agencies. To handle this complex data set, the IDNR is using the modernized version of the Environmental Protection Agency (EPA) database called STORET (STOrage and RETrieval). Not only does STORET allow the DNR to comprehensively handle its water monitoring data (including biological, physical, and chemical parameters), the use of STORET also ensures compatibility with data from other state and federal agencies.

It is envisioned that Iowa's STORET will be a "one-stop shopping" location for all the state's water monitoring data. This would include the coordination of data from other entities such as municipalities, citizen monitoring groups, universities and others into the STORET data management system. This data management system would serve both a warehousing function and an active data management system for the various agencies needing to access data on a regular basis. The warehouse streamlines access to data by removing the need to request data from each agency separately and provides the data in a standard format with the metadata intact. This is appealing to the public since they often do not know where to go to get the desired data. The use of STORET as a data management tool is currently in transition as the EPA migrates to a new, modern STORET database. At this point in time, STORET is functionally two separate databases: Legacy STORET and Modern STORET.

Legacy STORET

The original STORET database was developed 30 years ago to allow various agencies to share water quality information. STORET was a mainframe computer database operated and maintained by the EPA. Government users of the database were assigned user ids and passwords to store and retrieve data. Public access to the data was only available through the EPA and required a staff member to query the database and send the results to the requestor. Legacy STORET housed chemical, physical and biological data, but did not provide information on the monitoring metadata. Metadata is the "data about the data" and contains information such as the sample equipment, sampling methodology, laboratory techniques and the accuracy of measurements such as location, elevation and analytical results. This type of information is important because it provides quality control of the data. As computer technology has advanced and the volume of monitoring activities has increased, it became necessary to overhaul the STORET system into a

more flexible, modern system. This modernization began in the early 1990s and has resulted in a new STORET database. However, much of the previous STORET data does not contain information required in the new system. Therefore the old or "legacy" data has been migrated to a separate Oracle™ database and is archived for the EPA at the Legacy Data Center (LDC) in Research Triangle Park, North Carolina. The Legacy STORET database contains information through 1998, but was "frozen" January 1, 1999. All data collected after January 1, 1999 is housed in Modern STORET. Public access to the LDC is available through the EPA web page. States may migrate legacy data to the new system once the data has been documented and the proper metadata requirements have been fulfilled. The IDNR intends to migrate much of its legacy data to the new, modern system.

Modern STORET

Modern STORET was designed using the concept of a distributed or decentralized data management system. Advances in PC and client/server technology have revolutionized the way data is handled and made the use of mainframe computer systems less necessary. By abandoning the mainframe or centralized database, the modern system allows each agency to install STORET locally and customize the database to meet its specific management needs. Modern STORET runs on a powerful Oracle™ database platform and is therefore relational to other databases used to manage water resources. In this way, Iowa STORET data will be easily shared and integrated with data from other states. Modern STORET improves upon the Legacy system by storing metadata information and, by keeping data ownership at a local level, providing a more flexible database to users. Data are managed and stored at the local level and agencies may upload their STORET information to the centralized EPA data warehouse. In contrast to the old system where only the EPA was able to make changes to the mainframe system, the data warehouse serves only as a repository. The original and working copy of the database remains at the local level. Iowa is utilizing this data management structure to coordinate water-monitoring data from other agencies in the state. These agencies are assigned a unique organizational identification number, which establishes data ownership. IDNR works with these agencies to develop conduits for the data to flow from the original source database (whether it be Oracle, Microsoft Access, SQL server, etc.), translate the data into appropriate STORET formats, and then populate the STORET tables for that organization. While the system currently requires IDNR to handle the data, the future plans for the clearinghouse will allow other organizations to directly upload data through Web-based applications, removing the need for direct IDNR staff involvement.

Data Retrieval

Data warehousing or storage is only the first step in providing one-stop access to Iowa's water quality data. Providing the data in a cost and time efficient manner requires the use of tools widely available to public and private entities, and minimizing the amount of staff time devoted to handling data requests was essential. The IDNR has been serving data in the form of Geographic Information System (GIS) coverages on the Internet since the early 1990s and was moving in the direction of interactive mapping of data using ArcIMS technology. A web-based approach to serving the water quality data seemed the most logical choice given the widespread access to the Internet (every public school and library in the state has Internet access). The use of the Internet as a front-end access point for the data allows users to retrieve data without actually accessing the database, thus providing security for the clearinghouse. This allows IDNR to restrict access to areas of the database or block the viewing of data from a particular organization. The clearinghouse web page was released in April of 2001 and is limited in the amount of data currently available (only IDNR data collected as part of the ambient water monitoring program is available). Currently, water quality data can be downloaded from the Water Monitoring Program's web page at **wqm.igsb.uiowa.edu/STORET**. The interface allows a user to search by site, county, hydrologic unit (HUC), parameter, or month. Data can be downloaded as a flat-file or as a Microsoft Excel file. ArcIMS technology allows the user to display the monitoring sites and use standard GIS functions to query the data. This allows a user to click on a station or set of stations to view the data, display the data

in graphical form, or see statistical summaries. Through time, both the amount of data and the web page interface will be updated to enhance the clearinghouse functions. Plans also call for live, interactive graphing, charting, mapping and statistical summaries that provide information to the user along with the ability to download raw data.